

EXHIBIT G

recordings. In that case, Google was permitted to scan, digitize and catalog books in an online database after arguing that it had reproduced only snippets of the works online and had transformed the originals, which made it fair use.

Using data to train A.I. systems, Meta's lawyers said in their meetings, should similarly be fair use. At least two employees raised concerns about using intellectual property and not paying authors and other artists fairly or at all, according to the recordings. One employee recounted a separate discussion about copyrighted data with senior executives including Chris Cox, Meta's chief product officer, and said no one in that meeting considered the ethics of using people's creative works.

'Synthetic' Data

OpenAI's Mr. Altman had a plan to deal with the looming data shortage.

Companies like his, he said at the May conference, would eventually train their A.I. on text generated by A.I. — otherwise known as synthetic data.

Since an A.I. model can produce humanlike text, Mr. Altman and others have argued, the systems can create additional data to develop better versions of themselves. This would help developers build increasingly powerful technology and reduce their dependence on copyrighted data.

"As long as you can get over the synthetic data event horizon, where the model is smart enough to make good synthetic data, everything will be fine," Mr. Altman said.

A.I. researchers have explored synthetic data for years. But building an A.I. system that can train itself is easier said than done. A.I. models that learn from their own outputs can get caught in a loop where they reinforce their own quirks, mistakes and limitations.

"The data these systems need is like a path through the jungle," said Jeff Clune, a former OpenAI researcher who now teaches computer science at the University of British Columbia. "If they only train on synthetic data, they can get lost in the jungle."

To combat this, OpenAI and others are investigating how two different A.I. models might work together to generate synthetic data that is more useful and reliable. One system produces the data, while a second judges the information to separate the good from the bad. Researchers are divided on whether this method will work.

A.I. executives are barreling ahead nonetheless.

"It should be all right," Mr. Altman said at the conference.